SEQ ID NO:1. Human Hepsin DNA Sequence: The GenBank Accession No. for human hepsin mRNA (cDNA) is M18930:

	1	TCGAGCCCGC	TTTCCAGGGA	CCCTACCTGA	GGGCCCACAG	GTGAGGCAGC	CTGGCCTAGC
5	61	AGGCCCCACG	CCACCGCCTC	TGCCTCCAGG	CCGCCCGCTG	CTGCGGGGCC	ACCATGCTCC
•	121	TGCCCAGGCC	TGGAGACTGA	CCCGACCCCG	GCACTACCTC	GAGGCTCCGC	CCCCACCTGC
	181	TGGACCCCAG	GGTCCCACCC	TGGCCCAGGA	GGTCAGCCAG	GGAATCATTA	ACAAGAGGCA
	241	GTGACATGGC	GCAGAAGGAG	GGTGGCCGGA	CTGTGCCATG	CTGCTCCAGA	CCCAAGGTGG
	301	CAGCTCTCAC	TGCGGGGACC	CTGCTACTTC	TGACAGCCAT	CGGGGCGCA	TCCTGGGCCA
10	361	TTGTGGCTGT	TCTCCTCAGG	AGTGACCAGG	AGCCGCTGTA	CCCAGTGCAG	GTCAGCTCTG
	421	CGGACGCTCG	GCTCATGGTC	TTTGACAAGA	CGGAAGGGAC	GTGGCGGCTG	CTGTGCTCCT
	481	CGCGCTCCAA	CGCCAGGGTA	GCCGGACTCA	GCTGCGAGGA	GATGGGCTTC	CTCAGGGCAC
	541	TGACCCACTC	CGAGCTGGAC	GTGCGAACGG	CGGGCGCCAA	TGGCACGTCG	GGCTTCTTCT
	601	GTGTGGACGA	GGGGAGGCTG	CCCCACACCC	AGAGGCTGCT	GGAGGTCATC	TCCGTGTGTG
15	661	ATTGCCCCAG	AGGCCGTTTC	TTGGCCGCCA	TCTGCCAAGA	CTGTGGCCGC	AGGAAGCTGC
	721	CCGTGGACCG	CATCGTGGGA	GGCCGGGACA	CCAGCTTGGG	CCGGTGGCCG	TGGCAAGTCA
	781	GCCTTCGCTA	TGATGGAGCA	CACCTCTGTG	GGGGATCCCT	GCTCTCCGGG	GACTGGGTGC
	841	TGACAGCCGC	CCACTGCTTC	CCGGAGCGGA	ACCGGGTCCT	GTCCCGATGG	CGAGTGTTTG
	901	CCGGTGCCGT	GGCCCAGGCC	TCTCCCCACG	GTCTGCAGCT	GGGGGTGCAG	GCTGTGGTCT
20	961	ACCACGGGGG	CTATCTTCCC	TTTCGGGACC	CCAACAGCGA	GGAGAACAGC	AACGATATTG
	1021	CCCTGGTCCA	CCTCTCCAGT	CCCCTGCCCC	TCACAGAATA	CATCCAGCCT	GTGTGCCTCC
	1081	CAGCTGCCGG	CCAGGCCCTG	GTGGATGGCA	AGATCTGTAC	CGTGACGGGC	TGGGGCAACA
	1141	CGCAGTACTA	TGGCCAACAG	GCCGGGGTAC	TCCAGGAGGC	TCGAGTCCCC	ATAATCAGCA
	1201	ATGATGTCTG	CAATGGCGCT	GACTTCTATG	GAAACCAGAT	CAAGCCCAAG	ATGTTCTGTG
25	1261	CTGGCTACCC	CGAGGGTGGC	ATTGATGCCT	GCCAGGGCGA	CAGCGGTGGT	CCCTTTGTGT
	1321	GTGAGGACAG	CATCTCTCGG	ACGCCACGTT	GGCGGCTGTG	TGGCATTGTG	AGTTGGGGCA
	1381	CTGGCTGTGC	CCTGGCCCAG	AAGCCAGGCG	TCTACACCAA	AGTCAGTGAC	TTCCGGGAGT
	1441	GGATCTTCCA	GGCCATAAAG	ACTCACTCCG	AAGCCAGCGG	CATGGTGACC	CAGCTCTGAC
	1501	CGGTGGCTTC	TCGCTGCGCA	GCCTCCAGGG	CCCGAGGTGA	TCCCGGTGGT	GGGATCCACG
30	1561	CTGGGCCGAG	GATGGGACGT	TTTTCTTCTT	GGGCCCGGTC	CACAGGTCCA	AGGACACCCT
	1621	CCCTCCAGGG	TCCTCTCTTC	CACAGTGGCG	GGCCCACTCA	GCCCCGAGAC	CACCCAACCT
	1681	CACCCTCCTG	ACCCCCATGT	AAATATTGTT	CTGCTGTCTG	GGACTCCTGT	CTAGGTGCCC
	1741	CTGATGATGG	GATGCTCTTT	AAATAATAAA	GATGGTTTTG	ATT	
	25	5 61 121 181 241 301 10 361 421 481 541 601 15 661 721 781 841 901 20 961 1021 1081 1141 1201 25 1261 1321 1381 1441 1501 30 1561 1621 1681	5 61 AGGCCCACG 121 TGCCCAGGCC 181 TGGACCCCAG 241 GTGACATGGC 301 CAGCTCTCAC 301 TTGTGGCTGT 421 CGGACGCTCG 481 CGCGCTCCAA 541 TGACCACTC 601 GTGTGGACGA 721 CCGTGGACCG 721 CCGTGGACCG 781 GCCTTCGCTA 841 TGACAGCCGC 901 CCGGTGCCGT 20 961 ACCACGGGGG 1021 CCCTGGTCCA 1081 CAGCTGCCG 1141 CGCAGTACTA 1201 ATGATGTCTG 25 1261 CTGGCTACC 1321 GTGAGGACAG 1381 CTGGCTGTCC 1441 GGATCTTCCA 1501 CGGTGCCTT 30 1561 CTGGGCCGAG 1621 CCCTCCAGGG 1621 CCCTCCAGGG 1621 CCCTCCAGGG 1621 CCCTCCAGGG	5 61 AGGCCCACG CCACCGCCTC 121 TGCCCAGGCC TGGAGACTGA 181 TGGACCCCAG GGTCCCACCC 241 GTGACATGGC GCAGAAGGAG 301 CAGCTCTCAC TGCGGGGACC 10 361 TTGTGGCTGT TCTCCTCAGG 421 CGGACGCTCG GCTCATGGTC 481 CGCGCTCCAA CGCCAGGGTA 541 TGACCCACTC CGAGCTGGAC 601 GTGTGGACGA GGGGAGGCTG 601 GTGTGGACGA GGGGAGGCTG 721 CCGTGGACCG CATCGTGGAC 721 CCGTGGACCG CATCGTGGGA 781 GCCTTCGCTA TGATGGAGCA 841 TGACAGCCGC CCACTGCTTC 901 CCGGTGCCGT GGCCCAGGCC 20 961 ACCACGGGGG CTATCTTCCC 1021 CCCTGGTCA CCTCTCCAGT 1081 CAGCTGCCG CCAGGCCCTG 1141 CGCAGTACTA TGGCCAACAG 1201 ATGATGTCTG CAATGGCGCT 1321 GTGAGGACAG CATCTTCGG 1321 GTGAGGACAG CATCTCTCGG 1331 CTGGCTGTCC CCTGGCCCAG 1441 GGATCTTCCA GGCCATAAAG 1501 CGGTGGCTTC TCGCTGCCA 1441 GGATCTTCCA GGCCATAAAG 1501 CGGTGGCTTC TCGCTGCCAG 1621 CCCTCCAGGG TCCTCTCTTC 1681 CACCCTCCTG ACCCCCATGT	5 61 AGGCCCCACG CCACCGCCTC TGCCTCCAGG 121 TGCCCAGGCC TGGAGACTGA CCCGACCCCG 181 TGGACCCCAG GGTCCCACCC TGGCCCAGGA 241 GTGACATGGC GCAGAAGGAG GGTGGCCGGA 301 CAGCTCTCAC TGCGGGGACC CTGCTACTTC 10 361 TTGTGGCTGT TCTCCTCAGG AGTGACCAGG 421 CGGACGCTCG GCTCATGGTC TTTGACAAGA 481 CGCGCTCCAA CGCCAGGGTA GCCGGACTCA 541 TGACCCACTC CGAGCTGGAC GTGCGAACGG 601 GTGTGGACGA GGGGAGGCTG CCCCACACCC 721 CCGTGGACCG CATCGTGGAC GTGCGAACGG 721 CCGTGGACCG CATCGTGGAC GGCCGGACCA 781 GCCTTCGCTA TGATGGAGCA CACCTCTGTG 841 TGACAGCCGC CACTGCTTC CCGGAGCGGA 781 GCCTTCGCTA TGATGGAGCA CACCTCTGTG 841 TGACAGCCGC CCACTGCTTC CCGGAGCGGA 781 CCGTGCCGT GGCCCAGGCC TCTCCCCACG 1021 CCCTGGTCCA CCTCTCCAGT CCCCTGCCCC 1081 CAGCTGCCGG CCAGGCC TCTCCCCACGC 1021 CCCTGGTCCA CCTCTCCAGT CCCCTGCCCC 1081 CAGCTGCCGG CCAGGCCT GTGGATGGCA 1141 CGCAGTACTA TGGCCAACAG GCCGGGGTAC 1201 ATGATGTCTG CAATGGCGCT GACTTCTATG 1321 GTGAGGACAG CATCTCTCGG ACGCCACGTT 1331 CTGGCTGTGC CCTGGCCCAG AAGCCAGGCG 1441 GGATCTTCCA GGCCATAAAG ACTCACTCCG 1501 CGGTGGCTTC TCGCTGCCCA GCCTCCAGGG 30 1561 CTGGGCCGAG GATGGACGT TTTTCTTTT 1621 CCCTCCAGGG TCCTCTCTTC CACAGTGGCG 1681 CACCCTCCTG ACCCCCATGT AAATATTGTT	5 61 AGGCCCACG CCACCGCCTC TGCCTCAGG CCGCCGCTG 121 TGCCCAGGCC TGGAGACTGA CCCGACCCG GCACTACCTC 181 TGGACCCAG GGTCCCACCC TGGCCCAGGA GGTCAGCCAG 241 GTGACATGGC GCAGAAGGAG GGTGGCCGAA CTGTGCCATG 301 CAGCTCTCAC TGCGGGGACC CTGTCACTT TGACAGCCAT 10 361 TTGTGGCTGT TCTCCTCAGG AGTGACCAGG AGCCGCTGTA 421 CGGACGCTCG GCTCATGGTC TTTGACAAGA CGGAAGGAG 481 CGCGCTCCAA CGCCAGGGTA GCCGGACTCA GCTGCGAGGA 541 TGACCCACTC CGAGCTGGAC GTGCGAACGG CGGGCGCCAA 601 GTGTGGACGA GGGGAGCTG CCCCACACCC AGAGGCCCTA 601 GTGTGGACGA GGGGAGGCTG CCCCACACCC AGAGGCTGCT 15 661 ATTGCCCCAG AGGCCGTTTC TTGGCCGCCA TCTGCCAAGA 721 CCGTGGACCG CATCGTGGA GGCCGGACAC CCAGCTTGGG 781 GCCTTCGCTA TGATGGAGCA CACCTCTGTG GGGGATCCT 841 TGACAGCCGC CACTGCTTC CCGGAACGG GTCTGCCAAGA 721 CCGTGGACCG CATCGTGGA GGCCGGACA ACCGGCTCT 841 TGACAGCCGC CACTGCTTC CCGGAACGG GTCTGCCCCT 901 CCGGTGCCT GGCCCAGCC TCTCCCACG GTCTGCCCC 1021 CCCTGGTCCA CCTCTCCCAGT CCCCACG GTCTGCGCC 1021 CCCTGGTCCA CCTCTCCAGT CCCCACG GTCTGCCCC TCACAGCGA 1021 CCCTGGTCCA CCTCTCCAGT CCCCACG GTCTGACAGATA 1081 CAGCTGCCG CCAGGCCCTG GTGGATGGCA AGATCTGTAC 1141 CGCAGTACTA TGGCCAACAG GCCGGGGTAC TCCAGAGATA 1081 CAGCTGCCG CCAGGCCCTG GTGGATGGCA AGATCTGTAC 1201 ATGATGTCT CAATGGCGCT GACTTCTATG GAAACCAGAT 1201 ATGATGTCT CAATGGCGCT GACTTCTATG GAAACCAGAT 1201 ATGATGTCT CAATGGCGCT GACTTCTATG GAAACCAGAT 1321 GTGAGGACAG CATCTCTCGG ACCCCAGGCC TCCAGGGCGA 1321 GTGAGGACAG CATCTCTCGG ACCCCAGGCC TCCAGGGCGA 1321 GTGAGGACAG CATCTCTCGG ACCCCAGGCC TCCAGGGCAGA 1321 GTGAGGACAG CATCTCTCGG ACCCCAGGCC TCCAGGGCAGA 1441 GGATCTTCCA GGCCATAAAG ACTCACTCG AAGCCAGCGA 1441 GGATCTTCCA GGCCATAAAG ACTCACTCG AAGCCAGCGA 1441 GGATCTTCCA GGCCATAAAAG ACTCACTCG AAGCCAGCGA 1441 GGATCTTCCA GGCCATAAAAG ACTCACTCG AAGCCAGCGG 1501 CCGGGGCCAG GGCCATCAGG GGCCCACTCA 1681 CACCCTCCTG GATGGACG TTTTTCTTTT GGGCCCAGCGG 1501 CCGGGCCGAG GATGGACGT TTTTCTTTT GGGCCCAGCGG 1621 CCCTCCAGGG GATGGACGT TTTTCTTTT GGGCCCAGCTG 1681 CACCCTCCTG ACCCCCATGT AAATATGTT CTGCTGTCTG	121 TGCCCAGGCC TGGAGACTGA CCCGACCCCG GCACTACCTC GAGGCTCCGC 181 TGGACCCCAG GGTCCCACCC TGGCCCAGGA GGTCAGCCAG GGAATCATTA 241 GTGACATGGC GCAGAAGGAG GGTGGCCGA CTGTGCCATG CTGCTCCAGA 301 CAGCTCTCAC TGCGGGGACC CTGCTACTTC TGACAGCCAT CGGGGCGGCA 421 CGGACGCTCG GCTCATCGTC TTGACAGCCAT CCCAGTGCAG 421 CGGACGCTCG GCTCATGGTC TTTGACAAGA CGGAAGGGAC GTGGCGCTG 481 CGCGCTCCAA CGCCAGGGTA GCCGGACTCA GCTGCAGGAC GATGGCCTTC 541 TGACCCACTC CGAGCTGGAC GTGCGAACAGG CGGAAGGGAC GTGGCGCTG 601 GTGTGGACGA GGGGAGGCTG CCCCCACACCC AGAGGCTCT GGAGGTCATC 601 GTGTGGACGA GGGGAGGCTG CCCCACACCC AGAGGCTCT GGAGGTCATC 721 CCGTGGACCG CATCGTGGGA GGCCGGACA CCAGCTTGGG 721 CCGTGGCCG CATCGTGGGA GGCCGGACA CCAGCTTGGG 721 CCGTGGCCG CACCTGCTC CCGGAGCGAA ACCGGTTCCT GCCGAGGAC 721 CCGTGGCCG CACCTGCTC CCGGAGCGAA ACCGGGTCCT GTCCCGAGG 721 CCGTGCCG CCACTGCTC CCGGAGCGAA ACCGGGTCCT GTCCCGAGG 721 CCGGTGCCG CCACTGCTC CCGGAGCGAA ACCGGGTCCT GTCCCGATGG 721 CCGGTGCCG CCACTGCTC CCCGAGCCC TCCCCAACACACAG GGAGAACAGC 721 CCGGTGCCG CCACTGCTC CCCGGAGCGAA ACCGGGTCCT GTCCCGATGG 721 CCGGTGCCG CCACTGCTC CCCTGCCC CCAACACACAG GGAGAACAGC 721 CCGGTGCCG CCAGGCCCT TTCCCCACG GTCTGCAGCT GCGGGTGCAG 722 CCTGGCCCA CCTTCCCAGG CCCAACACAGAA CACCGGTGCAA ACCGGGTCAA CACCCTGCCC CCAACACAGAA CACCCCTG CCCACACACAGAA CACCCCAGGCC TCCCCACCCC TCCCACACACAGAA CACCCCAGGCC TCCCCACACACACAACACA

TU

35

SEQ ID NO:2. Human Hepsin Polypeptide Sequence (417 a.a.): The protein_id number is "AAA36013.1:

```
40

NH<sub>2</sub>-

1 MAQKEGGRTV PCCSRPKVAA LTAGTLLLT AIGAASWAIV AVLLRSDQEP LYPVQVSSAD
61 ARLMVFDKTE GTWRLLCSSR SNARVAGLSC EEMGFLRALT HSELDVRTAG ANGTSGFFCV
121 DEGRLPHTQR LLEVISVCDC PRGRFLAAIC QDCGRRKLPV DRIVGGRDTS LGRWPWQVSL
181 RYDGAHLCGG SLLSGDWVLT AAHCFPERNR VLSRWRVFAG AVAQASPHGL QLGVQAVVYH
241 GGYLPFRDPN SEENSNDIAL VHLSSPLPLT EYIQPVCLPA AGQALVDGKI CTVTGWGNTQ
301 YYGQQAGVLQ EARVPIISND VCNGADFYGN QIKPKMFCAG YPEGGIDACQ GDSGGPFVCE
361 DSISRTPRWR LCGIVSWGTG CALAQKPGVY TKVSDFREWI FQAIKTHSEA SGMVTQL
-COOH
```